ORIGINAL ARTICLE

EUROPEAN FINANCIAL MANAGEMENT WILEY

Using accounting-based and loan-related information to estimate the cure probability of a defaulted company

Christian Lohmann[1,2] | Thorsten Ohliger[3]

[1]Schumpeter School of Business and Economics, University of Wuppertal, Germany

[2]Faculty for Social Sciences, Economics, and Business Administration, University of Bamberg, Germany

[3]parcIT GmbH, Cologne, Germany

Correspondence
Christian Lohmann, Schumpeter School of Business and Economics, University of Wuppertal, Gaußstraße 20, 42119 Wuppertal, Germany.
Email: lohmann@wiwi.uni-wuppertal.de

Abstract

The cure of a defaulted company has important implications for the estimation of the loss given default. In this study, we estimate the probability of a defaulted company being cured using data on a large international sample of defaulted companies. More specifically, we examine whether historic accounting information on a defaulted company and loan-related information are associated with that company's probability of being cured. The main finding of our analysis is that both accounting-based and loan-related independent variables increase the validity of cure prediction models.

KEYWORDS
accounting information, company cure, collateralization, cure probability, defaulted company, Global Credit Data

JEL CLASSIFICATION
C53; D81; G24; G33

- -

1 | INTRODUCTION

The loss given default (LGD) values of companies that default and are subsequently cured tend to be low. As cured companies resume repaying both their loan and the interest on that loan, the LGD values they exhibit are frequently equal to or close to 0. This indicates that a defaulted company's probability of being cured greatly affects the LGD. Consequently, accurately estimating a company's probability of being cured could help improve estimations of the LGD.

The LGD distribution usually exhibits a bimodal structure that reaches the maximum number of observations when LGD values are close to 0 and when they are close to 1. The value of the LGD is close to 0 if either a defaulted company is cured or the collaterals and securities sufficiently cover the amount outstanding. Estimating a defaulted company's probability of being cured would therefore allow us to estimate the LGD values more accurately, considering that cure is associated with specific low LGD values. Estimating the LGD on the basis of historic data commonly involves taking into account observations of cured companies with LGD = 0 (Calabrese & Zenga, 2010; Renault & Scaillet, 2004). However, incorporating the potential cure event of a defaulted company into the LGD estimation requires a valid estimation of the probability that a defaulted company will be cured.

In the literature, there is no consensus as to whether a company's probability of being cured is associated with the estimation of LGD or with the estimation of the probability of default (PD). The case where the cure probability has to be taken into account by estimating LGD is in line with the regulatory specification that the cure event does not affect the assessment of the previous default event. For example, the Capital Requirement Regulation (CRR) of the European Union provides a rigorous definition of the default event in article 178 CRR (European Banking Authority, 2016), which, however, does not distinguish between soft and hard defaults. In the second case, the probability of a company being cured is taken into account when estimating the PD. This estimation requires both a default model that takes into account all defaulted companies, irrespective of whether they are subsequently cured, and a cure model to estimate a defaulted company's probability of being cured. Such estimations may involve a mixture cure model; that is, a special type of survival model that incorporates the possibility of cure (Beran & Djaidja, 2007; Dirick, Claeskens, & Baesens, 2015, 2017; Mo & Yau, 2010; Tong, Mues, & Thomas, 2012; Yildirim, 2008; Zhang, Yang, Kelleher, & Si, 2019). Alternatively, they may involve a cure-after-default model, which combines an upstream cure model with a subsequent default model (Wolter & Rösch, 2014).

Rigorous definitions of default allow us to estimate additionally a defaulted company's probability of being cured (PC), which improves estimations of the LGD. When the PC is accurately estimated, it is possible to separate the LGD into two distinct components. The first component relates to the expected LGD in the case of defaulted companies that are not subsequently cured. This component can be calculated by multiplying the downside probability $(1 - PC)$ and the LGD given a company's non-cure ($LGD_{NonCure}$). The second component represents the expected LGD in the case of defaulted companies that are subsequently cured; this component can be calculated by multiplying the upside probability PC and the LGD given a company's cure (LGD_{Cure}). Thus the LGD consists of two components:

$$LGD = (1 - PC) \cdot LGD_{NonCure} + PC \cdot LGD_{Cure}. \tag{1}$$

Breaking down the estimation of the LGD into $LGD_{NonCure}$ and LGD_{Cure} reduces the bimodal nature of the LGD distribution, because cured companies largely explain why the LGD is often 0, although there are no sufficient collaterals or securities, if any. The LGD_{Cure} component often takes a value close to 0 and also exhibits very low scatter. In comparison, $LGD_{NonCure}$

exhibits a less pronounced bimodal distribution, allowing $LGD_{NonCure}$ to be estimated more precisely. However, breaking down the LGD estimation as we have described requires that a company's probability of being cured is estimated as accurately as possible and that a financial institution that estimates the LGD is well informed about which parameters are associated with high or low cure probabilities.

To date, a defaulted company's probability of being cured has been empirically examined only by Wolter and Rösch (2014), who analyzed a sample of German firms on the basis of data covering the period 2002–2007. Wolter and Rösch (2014) found that the probability of cure increases with logarithmic sales and decreases with the financial ratios of current and long-term liabilities to total capital and of long-term provisions to total capital, as well as the logarithmic total capital. There are further studies on estimating the PD (see overviews by, for example, Altman & Saunders, 1998; Balcaen & Ooghe, 2006; Beaver, Correia, & McNichols, 2010; Bellovary, Giacomino, & Akers, 2007) and the LGD (see, for example, Bastos, 2010; Hagmann, Renault, & Scaillet, 2005; Loterman, Brown, Mertens, Mues, & Baesens, 2012); however, these studies do not focus on estimating the probability of cure.

The present study estimates the probability of a defaulted company being cured on the basis of data drawn from a large international sample of defaulted companies and analyzes the relationship between a set of independent variables and the probability of cure. In particular, the present study focuses on the explanatory power of accounting information on companies that were cured following their default. In addition, this analysis takes into account metric independent variables that relate to the credit transaction, as well as categorical variables that describe the indebted company in more detail. We estimate and examine in greater depth the relationship between these independent variables and a defaulted company's probability of being cured. For that purpose, we calculate a defaulted company's probability of being cured at 1 year before default. In order to determine the expected and unexpected loss accurately, it is necessary to be able to estimate a company's probability of being cured at 1 year before default. The estimated probability that a defaulted company will be cured is then part of estimating the LGD.

The present study provides insights into accurately calculating a defaulted company's probability of being cured and into the factors that influence this probability. Our main finding is that accounting-based independent variables describing the economic substance and the creditworthiness of a company increase the validity of models that predict this probability at least as much as loan-related independent variables do. More specifically, the findings of the present analysis provide clear empirical evidence that a company's probability of being cured decreases with the ratio of total loan volume to total assets, with the debt ratio, and with logarithmic sales. The relationship between the probability of cure and logarithmic sales contradicts the findings of Wolter and Rösch (2014) as they show that the probability of cure increases with logarithmic sales. Furthermore, our analysis shows that this probability decreases with the total loan volume, the drawn percentage in the lender limit, and the percentage of collaterals and securities in the outstanding amount. These findings indicate that both accounting-based information that describes a company's economic substance and creditworthiness and information that relates to the loan can help predict more accurately that company's probability of being cured. The comparison of validity measures in several estimations of this probability provides empirical evidence that accounting-based independent variables are at least as important as loan-related independent variables for accurate estimations of a company's probability of being cured. Consequently, both types of information should be taken into account in cure prediction models.

In the next section we describe the database from which we drew the data on which this analysis is based, as well as our approach to collecting and processing these data. We also

provide descriptive statistics of the refined sample. In Section 3, we present the results of several model estimations that predict a defaulted company's probability of being cured and we evaluate the validity of those models on the basis of several validity criteria. In Section 4 we test the robustness of our results. We conclude the paper in Section 5 with a summary of our main findings and discuss the explanatory power of accounting-based and loan-related information.

2 | EMPIRICAL DATA

2.1 | Sample refinement

The empirical analysis is based on real data on defaulted loans, which we collected from Global Credit Data, an international, non-profit association owned by its member banks. The raw data we collected on 5,325 defaulted companies contain information on the company, the loan, collaterals and securities, select accounting information from the annual financial statement, and the default and its settlement. Global Credit Data applies the definition of default that is based on Basel II regulations (e.g. Basel Committee on Banking Supervision, 2019, CRE36.69; Wagner, 2016). According to Global Credit Data. all member banks apply this definition of default. Our data also provide information on whether the defaulted company was cured and resumed the repayment of the loan. Our sample comprises only defaulted companies whose settlement process had been completed by December 31, 2018. For that reason, the information on a company's cure is binary coded. In our analysis, a defaulted company's cure is represented by the dependent variable. This takes the value $y_i = 1$ if the defaulted company i was cured and $y_i = 0$ if the defaulted company i was not cured.

Both the dependent variable, which represents a defaulted company's cure, and the independent variables have to be aggregated from the loan level to the company level, because the cure of a defaulted company is linked to all loans a company may have contracted. For that purpose, we first calculated the independent variables for each loan and then we aggregated the information on each individual loan that every company in our sample had. For example, this procedure relates to the calculation of the percentage of collaterals and securities in the total loan volume. We divided the large set of independent variables that we derived from the Global Credit Data database into subsets of metric variables that relate to the loan, the collaterals, and the company and of categorical variables that relate to the company. Table 1 provides an overview of the independent variables on the basis of scale and content.

The metric independent variables that relate to the loan include the total loan volume 1 year before default (TLV) and the drawn percentage in the lender limit 1 year before default (LIMIT). TLV reflects the outstanding credit 1 year before default, while LIMIT represents the ratio of the outstanding credit to the available credit limit 1 year before default. It is possible that LIMIT is not defined, if the available credit limit is equal to zero 1 year before default—for example, if a company defaults within the first year after being granted the credit. We eliminated from our database all observations where the metric independent variable LIMIT was undefined.

The metric independent variables reflecting collaterals include the percentage of collaterals and securities in the outstanding amount 1 year before default (COLL). The Global Credit Data database contains detailed information on the collaterals and other kinds of securities that are associated with several of the loans a defaulted company has. However, in the case of defaulted companies that have more than one outstanding loan, these data are not sufficiently detailed to allow us accurately to allocate these collaterals and securities to the different loans.

TABLE 1 Independent variables

This table reports the independent variables on the basis of scale and content.

Variable	Scale level	Content and categories
TLV	Metric	Total loan volume 1 year before default
LIMIT	Metric	Drawn percentage in the lender limit 1 year before default
COLL	Metric	Percentage of collaterals and securities in the outstanding amount 1 year before default
REL_TLV	Metric	TLV over total assets obtained from the balance sheet that was most recently disclosed between 2 years and 1 year before default
DEBT	Metric	Debt ratio obtained from the balance sheet that was most recently disclosed between 2 years and 1 year before default
LN(SALES)	Metric	Logarithmic value of sales obtained from the profit and loss statement that was most recently disclosed between 2 years and 1 year before default
IND	Nominal	Industry of the debtor (8 categories): (1) construction; (2) forestry, agriculture, and mining; (3) manufacturing; (4) real estate; (5) service and public sector; (6) trade; (7) transport and logistics; (8) utilities and telecoms
REGION	Nominal	Location of the debtor (5 categories): (1) eastern Europe, including CIS states; (2) North America; (3) miscellaneous; (4) southern and south-eastern Europe; (5) western and northern Europe

For the purposes of our analysis, we aggregated information on the collaterals and securities of each defaulted company in our sample. We then calculated the percentage of each outstanding loan that collaterals and other kinds of securities represent, taking into account real collaterals, such as mortgages, and personal securities, such as guarantees. We cut the percentage of an outstanding loan that is secured by collaterals and other kinds of securities at 100% because we had no information on the declaration of purpose. Although we lost some information as a result, we avoided creating the erroneous impression that all collaterals and securities serve the collateralization of all credit accounts, which is unlikely to be the case. In the final step of this procedure, we aggregated the percentages of outstanding loans relating to each defaulted company and derived the percentage of collaterals and securities at the company level.

We also derived the accounting-based independent variables from the Global Credit Data database due to the anonymity of the defaulted companies. Specifically, these sets of variables originated from each company's balance sheet and from the most recent profit and loss statement that had been issued between 2 years and 1 year before default. The variable REL_TLV represents the ratio of TLV to total assets, which we obtained from the balance sheet that we used. REL_TLV shows the relative meaning of the loans for the defaulted company. It should be noted that the disclosure date of the balance sheet does not always coincide with the date which is 1 year before default. The balance sheet we used was issued up to 12 months before that date which is 1 year before default. As a result, the accounting information of total assets does not include business activities that occurred between the balance-sheet date and the date which is 1 year before default.

We also used two additional accounting-based independent variables. First, we used the debt ratio (DEBT), which was derived from the last balance sheet each company published between 2 years and 1 year before default. Second, we used the logarithmic value of sales LN (SALES), which was derived from the last profit-and-loss statement each company published between 2 years and 1 year before default. We furthermore took into account the decreasing marginal effect of increasing sales and compressed the large value range of sales by taking the logarithm of the sales figure.

We included in our analysis the categorical independent variables industry (IND) and region (REGION). Although there are more granular data on IND and REGION, the categorical independent variable IND represents eight industries while REGION represents five regions (see Table 1). Each category includes more than 100 observations.

Macroeconomic conditions in downturn periods have to be taken into account when predicting the PD (e.g. Couderc, Renault, & Scaillet, 2008; Jones, 2017) and LGD (e.g. Calabrese, 2014; Krüger & Rösch, 2017). The macroeconomic conditions may also affect the probability of defaulted companies being cured as the cure rate varies in time and within the business cycle. The cure rate is relatively low in downturn years (2000–2001 and 2008–2009) and relatively high in boom phases (2005–2007 and 2013–2017); see the descriptive statistics in Table 6. To control for macroeconomic conditions in downturn periods we included the metric macroeconomic variable GDP that measures the growth in gross domestic product 2 years before default. The data on national annual GDP growth rates were obtained from the World Bank. We mapped GDP at country level. Where the annual GDP growth rate was not given for a specific year–country combination, we instead used the average annual GDP growth rate of all countries in the greater region (REGION) in which the country in question is located.

As already mentioned, to perform our analysis, we drew data on 5,325 defaulted companies from the Global Credit Data database. We only used complete data sets that contained all the information that our dependent and independent variables reflect. In Table 2 we describe the five steps of the procedure we followed in order to collect and process our data. Through this procedure,

TABLE 2 Processing the raw data

This table reports the five-step procedure of processing the raw data to refine the final sample.

	Number of defaulted companies	Number of cured companies
Collected detailed data on defaulted companies, derived from the Global Credit Data database	5,325	
(1) Eliminated state-owned, non-profit, and financial companies	5,257	
(2) Eliminated observations with infinite independent variables	4,855	1,779 (36.64%)
(3) Eliminated observations before January 1, 2000	4,486	1,750 (39.01%)
(4) Eliminated observations with outliers	4,073	1,641 (40.29%)
(5) Split observations into a training sample and a validation sample at a ratio of 2 to 1		
Training sample	2,715	1,109 (40.85%)
Validation sample	1,358	532 (39.18%)

–WILEY |

we extracted and refined the final sample that we used in our empirical analysis. We started by eliminating all state-owned, non-profit, and financial companies; this allowed us to obtain a homogeneous sample. We then cleaned up our sample, removing all observations with infinite metric independent variables LIMIT or REL_TLV. Subsequently, we also removed all companies that defaulted before January 1, 2000. In the next step we identified the outliers. In cases where the metric independent variable had no natural upper or lower limit (e.g. the natural lower limit of the debt ratio is 0), we identified the outliers using either the 2% or the 98% quantile or both. We then considered the advantages and disadvantages of either eliminating or adjusting the outliers (Baruch & Sunder, 1979). The advantage of trimming is that the resulting sample is unbiased; however, this approach reduces the sample size. In contrast, winsorizing compresses the range of the independent variables artificially, as outliers are set at the threshold, but preserves the sample intact. Having considered both approaches, we decided to trim our sample so as to avoid distorted estimations. This step reduced our data to 4,073 defaulted companies. Of those companies, 1,641 (40.29%) were finally cured and resumed the loan repayment process. Finally, we randomly split our sample of 4,073 observations into a training sample and a validation sample at a ratio of 2 to 1.

2.2 | Descriptive statistics

The descriptive statistics we present in Table 3 are based on the training sample of 2,715 defaulted companies. The mean of the metric variable LIMIT shows that, in the majority of cases, the defaulted company has almost exhausted its available credit limit. The data also show that in the majority of observations the metric independent variables TLV, COLL, and REL_TLV exhibit low values. The mean and the median of the metric variable COLL indicate that the majority of the defaulted companies have neither collaterals nor securities. In most cases REL_TLV is lower than DEBT, which includes both the TLV granted and other types of liabilities. However, TLV is likely to increase between the balance-sheet date and the date at 1 year before default because the total assets, which we derived from the last balance sheet that was released in the period between 2 years and 1 year before default, remain constant. With regard to geographical location and industry, most companies in our sample are located in Europe and are part of the construction, manufacturing, and trade sectors.

Comparison between the descriptive statistics of cured and non-cured companies shows that there are noticeable differences in the metric independent variables except for the metric independent variable LIMIT. Furthermore, Table 3 shows that the mean LGD of cured companies differs from that of non-cured companies. The mean LGD value of the cured companies is close to 0, whereas the mean LGD value of non-cured companies is significantly above 0. The LGD values of the defaulted companies were derived from the Global Credit Data database and the outliers were winsorized at the upper and lower limits of LGD = 0 and LGD = 2 respectively. The descriptive statistics demonstrate that cured and non-cured defaulted companies exhibit different LGD values and that accounting-based and loan-related independent variables might help accurately predict the probability of defaulted companies being cured.

2.3 | Correlations

For the purposes of our analysis, we use a multivariate generalized linear model with a logistic distribution function (e.g. Hosmer, Lemeshow, & Sturdivant, 2013; Maddala, 1984; Rauhmeier 2011).

TABLE 3 Descriptive statistics (training sample)

This table reports the descriptive statistics of the metric independent variables, the observed LGD, and the distributions of the categorical independent variables.

	All defaulted companies(2,715)			Cured companies(1,109)			Non-cured companies(1,606)		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
TLV	2,382,119	450,000	5,485,033	1,420,000	286,000	4,161,746	3,050,000	600,000	6,151,051
LIMIT	0.931	1.000	0.185	0.931	1.000	0.192	0.931	1.000	0.180
COLL	0.159	0.000	0.353	0.051	0.000	0.216	0.233	0.000	0.406
REL_TLV	0.230	0.094	0.406	0.194	0.076	0.350	0.255	0.106	0.440
DEBT	0.591	0.721	0.375	0.525	0.674	0.400	0.637	0.745	0.350
LN(SALES)	14.90	14.70	2.313	14.70	14.60	1.914	15.10	14.70	2.541
LGD	0.137	0.009	0.282	0.005	0.002	0.008	0.229	0.053	0.338
Construction			541	Eastern Europe, including CIS states				225	
Forestry, agriculture, and mining			106	North America				194	
Manufacturing			593	Miscellaneous				147	
Real estate			352	Southern and South-eastern Europe				1,370	
Service and public sector			309	Western and northern Europe				779	
Trade			572						
Transport and logistics			141						
Utilities and telecoms			101						

Our aim is to examine how select metric and categorical independent variables may affect a defaulted company's probability of being cured. This approach requires taking into account potential multi-collinearity between the independent variables. In a multivariate regression model, using correlated independent variables can make it hard to differentiate between the effects of the various independent variables on the dependent variable and may also influence the standard errors and the statistical significance tests of the corresponding estimated coefficients (Studenmund, 2016).

We used two measures to examine the correlations between each pair of independent variables on the basis of the lowest scale of two independent variables: the Bravais–Pearson correlation coefficient and Cramér's V. The former measures the correlation between metric independent variables; the latter measures the correlation between categorical independent variables and the correlation between a metric and a categorical independent variable, provided that the metric independent variable is classified into 10 categories based on quantiles. The correlations between each pair of independent variables, based on the training sample, are presented in Table 4.

Table 4 shows that a high correlation occurs between the metric independent variables TLV and LN(SALES). This is not surprising, given that both of these variables are indicative of a company's size. However, the value of the correlation, .40, does not affect the results of our analysis, but that correlation has to be taken into account by evaluating the estimations. Other correlations between the metric independent variables, between the metric and categorical independent variables, and between the categorical independent variables are low.

TABLE 4 Correlations (training sample)

This table reports the correlation coefficients according to Bravais–Pearson and Cramér's V, on the basis of the lowest scale level.

| Correlation coefficients according to Bravais–Pearson | | | | | | |
	TLV	LIMIT	COLL	REL_TLV	DEBT	LN(SALES)	GDP
TLV	1.00	−.03	.19	.06	−.03	.40	.01
LIMIT	−.03	1.00	−.07	.06	.25	−.22	−.10
COLL	.19	−.07	1.00	.17	−.07	.03	.16
REL_TLV	.06	.06	.17	1.00	−.00	−.31	.07
DEBT	−.03	.25	−.07	−.00	1.00	−.20	.11
LN(SALES)	.40	−.22	.03	−.31	−.20	1.00	−.04
GDP	.01	−.10	.16	.07	.11	−.04	1.00
Correlation coefficients according to Cramér's V						
	TLV	LIMIT	COLL	REL_TLV	DEBT	LN(SALES)	GDP
IND	.07	.09	.09	.12	.15	.15	.13
REGION	.06	.31	.24	.17	.43	.28	.53
	IND	REGION					
IND	1.00	.22					
REGION	.22	1.00					

The relationship between REGION and the other metric independent variables is more pronounced than it is between IND and the other metric independent variables. The high correlation between GDP and REGION is a plausible and expected observation. Overall, these correlations indicate that we do not need to apply further restrictions.

To compare the distributions of the independent variables between the defaulted companies that subsequently cured and those that were liquidated, we conducted a univariate analysis of the independent variables. With respect to the metric independent variables, the t-test and Pearson's median test show significant differences in the mean and median of most metric independent variables between these two groups of companies ($p < .001$). Only the difference in the mean of LIMIT ($p = .96$) and the difference in the median of LN(SALES) ($p = .21$) are not statistically significant. We also applied a chi-squared test to examine the distributions of the categorical independent variables. This test shows that there are statistically significant differences between the two groups of companies with regard to IND and REGION ($p < .001$).

3 | EMPIRICAL ESTIMATIONS

3.1 | Logistic models for predicting the probability of cure 1 year before default

In the logistic models we applied to predict a defaulted company's cure at 1 year before default we used the independent variables that capture this point in time. This point in time is commonly used in credit-risk management and to calculate the required own funds for a bank's credit risk (e.g. European Banking Authority, 2017, para. 122). We divided the final sample randomly into a training sample and a validation sample, using a ratio of 2 to 1. We then tested whether the results we obtained from applying the models on the training sample can be transferred to the validation sample. The validation analysis shows that they can.

We treated model 1a as a benchmark and used only the metric macroeconomic variable GDP and the categorical independent variables IND and REGION. In models 2a and 3a we also included loan-related independent variables TLV, LIMIT, and COLL, while in model 4a we added the accounting-based independent variables REL_TLV, DEBT, and LN(SALES). Finally, in models 5a and 6a we incorporated categorical, loan-related, and accounting-based independent variables. Due to the computational relationship between the independent variables TLV and REL_TLV, we estimated two models which separately take into account both independent variables. In models 5a and 6a we tested whether using both accounting-based and loan-related independent variables can increase the models' validity and predictive power. The regression analyses include the categorical independent variables IND and REGION; however, in Table 5 we only report the estimated coefficients of the metric independent variables and the respective levels of significance. We used a Wald test to examine the level of significance of each coefficient.

In model 1a, with only the metric macroeconomic variable GDP and the categorical independent variables IND and REGION, the estimated coefficient of GDP is negative and statistically significant. If the growth in GDP 2 years before default is high and, therefore, the company defaults during or directly after a good economic situation, a defaulted company's probability of being cured decreases. As the default is not caused by depressed market conditions during or directly after a downturn period, the default can be traced back to company-specific factors that may hinder a company's cure. The estimation also shows that the companies that are most likely to be cured after default are located in southern and south-eastern Europe and in western

TABLE 5   Model estimations

This table reports the estimated coefficients of the cure prediction models 1 year before default with regard to several combinations of the independent variables.

Variable	Model 1a	Model 2a	Model 3a	Model 4a	Model 5a	Model 6a
TLV		−4.37e−08 (.000)	−3.35e−08 (.002)		−1.39e−08 (.207)	
LIMIT		−1.028 (.000)	−0.840 (.002)		−0.905 (.001)	−0.906 (.001)
COLL			−1.748 (.000)		−1.616 (.000)	−1.614 (.000)
REL_TLV				−0.336 (.011)		−0.213 (.108)
DEBT				−2.032 (.000)	−1.866 (.000)	−1.874 (.000)
LN(SALES)				−0.081 (.001)	−0.071 (.005)	−0.094 (.000)
IND	Yes	Yes	Yes	Yes	Yes	Yes
REGION	Yes	Yes	Yes	Yes	Yes	Yes
GDP	−0.148 (.000)	−0.153 (.000)	−0.149 (.000)	−0.145 (.000)	−0.147 (.000)	−0.145 (.000)
Nagelkerke pseudo-R^2	.144	.161	.212	.221	.263	.263
Akaike information criterion	3,391.73	3,357.29	3,240.84	3,218.60	3,119.88	3,118.80
AUC (training)	.697	.708	.743	.745	.771	.771
AUC (validation)	.718	.728	.770	.767	.793	.793

and northern Europe. This probability is significantly lower in all other regions, particularly North America. The relationship between IND and a defaulted company's probability of being cured is unremarkable. One finding indicates that companies operating in the service and public sector are more likely to be cured after default than companies in the construction industry. All models produced similar results.

In model 2a the relationships between the metric independent variables TLV and LIMIT and a defaulted company's probability of being cured are negative and statistically significant. This probability decreases when the total loan volume is higher 1 year before default and when the ratio of outstanding credit to the available credit limit 1 year before default is higher. Considering that the independent variables TLV and LIMIT indicate a company's flexibility, we can conclude that a company with a more flexible financial structure is associated with a higher probability of being cured. However, the predictive power of the independent variable LIMIT is unexpected as there is no difference in the mean value of LIMIT between cured and non-cured defaulted companies. The reason for this statistically significant relationship is that we control

for REGION. The majority of observations with LIMIT ≥ 1 correspond to defaulted companies that are located in southern and south-eastern Europe. If we exclude defaulted companies that are located in southern and south-eastern Europe from the univariate analysis, we observe a difference in the mean value of LIMIT between cured and non-cured defaulted companies. We have taken this observation as a reason to check for multicollinearity by calculating the variance inflation factor (VIF). Model 2a and all other estimated models show VIF values below the threshold of VIF $= 5$, which indicates high multicollinearity (Sheather, 2009). The highest VIF is calculated for REGION in model 5a with VIF $= 4.72$.

Model 3a shows that the metric independent variable COLL has a statistically significant negative effect on a defaulted company's probability of being cured. Creditors will typically try to minimize the LGD by liquidating the collaterals and securities of a defaulted company. If a company has few or no collaterals and securities, the creditor has a greater incentive to attempt to restructure the defaulted company in order to minimize the LGD. In addition to this insight, it is worth noting that taking into account collaterals and securities increases a model's validity significantly, as the comparison between models 2a and 3a indicates.

In model 4a we added the accounting-based independent variables REL_TLV, DEBT, and LN (SALES). Higher values of the variable REL_TLV, which reflects the financial importance of the total loan volume for a company, and of DEBT, which indicates a company's financial health, are associated with a significantly lower probability of cure following default. This probability also decreases with a company's size, which is captured by LN(SALES). This finding contradicts the conclusion of Wolter and Rösch (2014) who empirically demonstrated that the probability of cure increases with logarithmic sales. Wolter and Rösch (2014) based their conclusion on a sample of German companies. Their explanation for their finding is that a slight reduction in cost might boost the margins and profitability of particularly large defaulted company and that it is easier to improve the market position of a large company. In contrast, our findings suggest that it may be easier to restructure a smaller rather than a larger company. The size effect is dominant in our sample and leads us to conclude that a defaulted company's probability of being cured decreases with LN(SALES).

The observed relationships between the independent variables and a defaulted company's probability of being cured are also present in the estimations of models 5a and 6a. Comparing model 5a with model 3a, we found that in model 5a the variable TLV does not have any statistically significant effect on a defaulted company's probability of being cured. The reason for this difference is that the independent variables TLV and LN(SALES), which reflect company size, are moderately correlated (.40). However, the independent variable REL_TLV, which is adjusted for company size, is also not statistically significant in model 6a as it is slightly above the significance level of 10%.

3.2 | Model validity

To compare the validity of the different models, we used Nagelkerke's pseudo-R^2, which is based on likelihood, and the area under the receiver operating characteristic curve (AUC), which is based on classification. Our analysis shows that both the independent variable COLL and the accounting-based independent variables have considerable explanatory power and impact on a model's validity. We tested for differences in AUC validity measures following DeLong, DeLong, and Clarke-Pearson (1988) and found that introducing additional independent variables significantly increases AUC values. However, we did not obtain significantly different AUC values when we compared model 3a

with model 4a and model 5a with model 6a. Comparing models 3a and 4a with regard to the AUC validity measures indicates that accounting-based independent variables contain a similar level of explanatory information to loan-related independent variables. The same comparison suggests that the role of accounting-based independent variables in predicting the probability of cure is at least as important as that of loan-related independent variables. Comparing models 5a and 6a with regard to the AUC validity measures shows that differentiating between TLV and REL_TLV and, therefore, normalizing company size does not substantially increase the accuracy of the prediction.

If all independent variables are incorporated in a cure prediction model, the validity measures reach their highest level. This applies to both the training and the validation sample and suggests that accounting-based independent variables can enhance the predictive power of models based exclusively on categorical and loan-related independent variables. Consequently, we recommend that accounting-based independent variables that describe the debtor in more detail should be taken into account in models predicting the probability of defaulted companies being cured.

3.3 | Application of the predicted probability of defaulted companies being cured to calculate the risk-weighted exposure

The predicted probability of defaulted companies being cured is a part of the LGD and therefore affects the calculation of the risk-weighted exposure that determines a financial institution's own funds according to the advanced internal ratings-based (A-IRB) approach. As an example, we analyze the risk-weighted exposure that is calculated with and without the predicted probability of defaulted companies being cured. If the difference in the risk-weighted exposure that is calculated with and without the predicted probability of defaulted companies being cured is substantial, the probability of defaulted companies being cured may be taken into account as a further risk parameter. To calculate the risk-weighted exposure (Regulation (EU) No. 575/2013,[1] Article 153), we have to make five simplifying assumptions:

1. The loans are taken into account as exposures to corporates (Regulation (EU) No. 575/2013, Article 147, para. 2(c)) and, in particular, as exposures to SMEs (Regulation (EU) No. 575/2013, Article 147, para. 5(a)(ii)).
2. We assume that the maturity of the loans is 2.5 years (Regulation (EU) No. 575/2013, Article 162, para. 1).
3. We assume that the loans are unconditionally cancelable at any time. This assumption allows the application of a conversion factor of 0% (Regulation (EU) No. 575/2013, Article 166, para. 8(a)).
4. We apply the internal estimated PDs that were derived from the Global Credit Data database. Each PD takes a value of at least 0.03% (Regulation (EU) No. 575/2013, Article 160, para. 1). Where there was no internal PD, we derived the PD from external ratings (S&P Global Ratings, 2019) instead. In that case, the PD corresponds to the PD that is derived from an internal ratings-based (IRB) approach. Due to missing PDs and implausible PDs (PD = 1), we were able to assign a PD to 1,527 companies (37.5% of the final sample of 4,073 companies).

5. We assume that the LGD of companies that are not cured is given by the percentage of the unsecured total loan volume. This means that the probability of defaulted companies being cured does not have any effect if the loan is completely collateralized. Furthermore, we assume that the LGD of companies that are cured is given by LGD = 0.

Based on these assumptions, we can calculate the risk-weighted exposure with and without the probability of defaulted companies being cured for 1,527 companies for which we have all the required information. For that purpose, we apply the probability of defaulted companies being cured that model 6a predicts. If we take into account the probability of defaulted companies being cured, the risk-weighted exposure decreases by 37%. Although this result is driven particularly by the assumptions on PD and LGD, the probability of defaulted companies being cured seems to have a substantial effect on the risk-weighted exposure that determines a financial institution's own funds.

4 | ROBUSTNESS OF THE EMPRICAL RESULTS

4.1 | Sample selection bias due to the duration of the settlement process

One concern we needed to address was that the sample selection might bias the results of our analysis. Our sample shows an increase in the rate of cure following default over time. This increase reflects the fact that our sample contains only defaulted companies whose settlement process had been completed by December 31, 2018 and that the settlement process is faster

TABLE 6 Annual distribution, 2000–2017

This table reports the annual distribution of cured and noncured companies following default.

Year of default	2000	2001	2002	2003	2004	2005
Defaulted companies	95	146	196	156	122	249
Non-cured companies	87	129	170	126	91	102
Cured companies	8	17	26	30	31	147
Percentage of cured companies	8.42%	11.64%	13.27%	19.23%	25.41%	59.04%
Year of default	2006	2007	2008	2009	2010	2011
Defaulted companies	248	359	311	619	317	282
Non-cured companies	105	114	208	506	184	147
Cured companies	143	245	103	113	133	135
Percentage of cured companies	57.66%	68.25%	33.12%	18.26%	41.96%	47.87%
Year of default	2012	2013	2014	2015	2016	2017
Defaulted companies	337	230	197	140	56	13
Non-cured companies	199	88	91	51	28	6
Cured companies	138	142	106	89	28	7
Percentage of cured companies	40.95%	61.74%	53.81%	63.57%	50.00%	53.85%

when a defaulted company is cured. The annual distribution of defaulted companies and the proportion of companies that were cured after default are displayed in Table 6 (full refined sample). The table shows that after 2012 the number of observations decreases as the number of defaulted companies whose settlement process had not been completed by December 31, 2018 increases. In this sample, the average cure rate is about 40%. However, the faster progress of the settlement process increases the proportion of cured companies: for example, in 2015 the proportion of defaulted and subsequently cured companies reaches 63.57%. To control for sample selection bias, we can exclude data from the years in which the faster settlement process distorts the overall sample composition.

To determine which observation period biases the sample composition, we calculated for each year the proportion of defaulted companies that had completed the settlement process by December 31, 2018. We can see that this proportion starts decreasing markedly after 2012. For example, only 55% of the companies that defaulted in 2017 had completed their settlement process by December 31, 2018. In contrast, 89% of the companies that defaulted in 2012 and 95% of the companies that had defaulted before 2012 had completed their settlement process by that date. On the basis of these figures, we grouped the companies that had defaulted by the end of 2012 into a separate subsample. We then repeated the five steps of processing our original raw data to obtain a second reassembled training sample consisting of 2,268 defaulted companies and 852 (37.57%) cured companies and a second reassembled validation sample of 1,135 defaulted companies and 410 (36.12%) cured companies. Using these samples, we repeated the correlation and multivariate regression analysis. Although the new, reduced sample is not biased by differences in the settlement periods of cured and non-cured companies, both analyses produce similar empirical results. The estimated coefficients of the metric independent variables and the related level of significance are displayed in Table 7. The level of significance of each coefficient was tested by means of the Wald test. The empirical results with regard to the metric independent variables exhibit similar estimated coefficients and similar levels of significance. Only the independent variable REL_TLV exhibits a higher level of significance as the corresponding p-value is $p > .006$ in model 6b.

The additional analysis shows that the selection of our sample does not distort the models we estimated to predict the probability of defaulted companies being cured. The validity measures in Table 7 are similar to the validity measures reported in Table 5. The ratios between the validity measures of all six models remain unchanged. When we compare the results of models 3b and 4b with the results of model 1b, we see that the loan-related independent variables and the accounting-based independent variables are informative to predict the probability of defaulted companies being cured. However, combining accounting-based and loan-related independent variables is the best way to increase considerably a model's validity, accuracy, and predictive power.

4.2 | Out-of-time validation

An out-of-time validation checks whether the relationships identified between the accounting-based and loan-related information and the probability of defaulted companies being cured persist in time and can therefore be used to predict the probability of cure for current or future loans. For that purpose, we reassembled the training and validation samples that we apply in models 1a–6a according to the year of default. The out-of-time training sample includes the 3,667 observations spanning the period 2000–2013 and corresponding to 2,256 non-cured and 1,411 cured companies. The validation sample contains the remaining 406 observations from the years 2014–2017 and corresponding to 176 non-cured and 230 cured companies.

TABLE 7 Model estimations (reassembled sample)

This table reports the estimated coefficients of the cure prediction models 1 year before default (reassembled sample).

Variable	Model 1b	Model 2b	Model 3b	Model 4b	Model 5b	Model 6b
TLV		−4.75e−08 (.000)	−3.93e−08 (.001)		−1.94e−08 (.132)	
LIMIT		−0.820 (.002)	−0.609 (.031)		−0.628 (.034)	−0.621 (.036)
COLL			−2.051 (.000)		−1.900 (.000)	−1.925 (.000)
REL_TLV				−0.474 (.005)		−0.462 (.006)
DEBT				−2.409 (.000)	−2.218 (.000)	−2.234 (.000)
LN(SALES)				−0.082 (.003)	−0.075 (.008)	−0.112 (.000)
IND	Yes	Yes	Yes	Yes	Yes	Yes
REGION	Yes	Yes	Yes	Yes	Yes	Yes
GDP	−0.004 (.000)	−0.005 (.000)	−0.005 (.000)	−0.004 (.000)	−0.004 (.000)	−0.004 (.000)
Nagelkerke pseudo-R^2	.121	.137	.209	.223	.274	.276
Akaike information criterion	2,812.57	2,788.10	2,651.67	2,625.28	2,524.28	2,519.03
AUC (training)	.678	.689	.741	.744	.772	.773
AUC (validation)	.692	.708	.751	.760	.795	.792

Using these reassembled samples, we repeated the model estimations, leading to the estimated models 1c–6c presented in Table 8. The level of significance of each coefficient was tested by means of the Wald test. The empirical results with regard to the metric independent variables exhibit similar estimated coefficients and similar levels of significance. Only the independent variable TLV now exhibits a low level of significance in model 5c, as the corresponding p-value is $p = .046$. The independent variable REL_TLV is statistically significant in model 6c as it is slightly below the significance level of 10%. Models 1c–6c confirm that combining accounting-based and loan-related independent variables can considerably increase a model's validity, accuracy, and predictive power. As a result, our empirical findings are confirmed by the out-of-time validation.

4.3 | Further robustness checks

The estimated models 1a–6a depend on the assumed distribution function. We applied the logistic distribution function in our estimations. The variance of the logistic distribution is

TABLE 8 Model estimations (out-of-time validation)

This table reports the estimated coefficients of the cure prediction models 1 year before default (out-of-time validation).

Variable	Model 1c	Model 2c	Model 3c	Model 4c	Model 5c	Model 6c
TLV		−5.26e−08 (.000)	−4.18e−08 (.000)		−2.09e−08 (.046)	
LIMIT		−1.020 (.000)	−0.871 (.000)		−0.885 (.000)	−0.902 (.000)
COLL			−2.100 (.000)		−1.982 (.000)	−1.993 (.000)
REL_TLV				−0.272 (.012)		−0.188 (.092)
DEBT				−2.415 (.000)	−2.238 (.000)	−2.256 (.000)
LN(SALES)				−0.088 (.000)	−0.078 (.000)	−0.104 (.000)
IND	Yes	Yes	Yes	Yes	Yes	Yes
REGION	Yes	Yes	Yes	Yes	Yes	Yes
GDP	−0.147 (.000)	−0.152 (.000)	−0.150 (.000)	−0.148 (.000)	−0.150 (.000)	−0.149 (.000)
Nagelkerke pseudo-R^2	.133	.153	.223	.232	.289	.289
Akaike information criterion	4,534.71	4,478.57	4,260.88	4,231.70	4,045.91	4,047.16
AUC (training)	.694	.705	.750	.753	.784	.783
AUC (validation)	.719	.725	.707	.749	.749	.747

greater than the variance of the standard normal distribution (Amemiya, 1981; Fahrmeir & Tutz, 2001). However, the choice of distribution functions should not significantly influence the results (Porath, 2006). To verify the effect of a different distribution function, we again estimated models 1a–6a by applying the standard normal distribution. The estimations were comparable to the estimations that are presented in Table 5.

Specific industry characteristics that are reflected in the accounting-based independent variables (Chava & Jarrow, 2004; Lev, 1969) may also affect the model estimations. To capture potential industry effects, we modified each accounting-based independent variable by calculating its relative deviation from the annual industry mean for each year (e.g. Berg, 2007; Lohmann & Ohliger, 2017). We then repeated the estimations of models 1a–6a by applying the modified accounting-based independent variables. As the estimations were comparable to the estimations that are presented in Table 5, we have to conclude that the results are not biased by specific industry characteristics.

The presence of influential observations may also distort the estimations. We identify influential observations in the training sample of 2,715 defaulted companies by applying Cook's distance (Cook, 1977) and the threshold $4/(n - k - 1)$, where n is the number of observations

und k is the number of independent variables (Hair, Anderson, Tatham, & Black, 1998). We estimated again models 1a–6a without the influential observations and obtained comparable results. Consequently, the relationships identified between the accounting-based and loan-related information and the probability of defaulted companies being cured are robust and not affected by influential observations.

Heteroskedasticity is another reason why the model estimations of the probability of defaulted companies being cured may not be robust. We control for heteroskedasticity by estimating heteroskedastic logit models (Efron, 1986; Nelder & Lee, 1998; Wilson & Lorenz, 2015, pp. 249–264). We applied a two-step procedure to estimate the heteroskedastic logit models. First, we identified those metric independent variables that have a statistically significant effect on the variance and, therefore, cause heteroskedasticity. Then we again estimated models 1a–6a as heteroskedastic logit models. In doing so, the identified metric independent variables were also used to estimate the latent scale models. The heteroskedastic logit models show comparable results and confirm the robustness of the estimations in Table 5.

5 | CONCLUSION

To predict the loss resulting from a company's default, it is vital accurately to estimate the debtor's PD, the exposure at default, and the LGD. The estimated LGD depends to a meaningful extent on the event that a defaulted debtor will be cured. The event of company cure thus reflects the associated non-performing loan becoming a performing loan again. The LGD of companies that are cured after default is usually very low and often close to 0. This suggests that the probability of cure affects the estimation of the LGD. In our study, estimating the probability of a company being cured following default allowed us to differentiate between cured and non-cured defaults in the LGD distribution of all observations. This distinction, in turn, helps increase the validity and accuracy of the LGD estimation, because the characteristic bimodal structure of the LGD is less pronounced in the separate LGD distributions.

The present study examines whether accounting-based and loan-related independent variables can help more accurately predict a company's probability of being cured following default. Our results show that using the independent variables we selected enables us to predict this probability with satisfactory accuracy. In particular, they show that the accounting-based independent variables we used and the independent variable reflecting collateralization have considerable explanatory power and impact on a model's validity. Our findings provide clear empirical evidence that a defaulted company's probability of being cured decreases with the ratio of total loan volume to total assets, debt ratio, and logarithmic sales. Furthermore, this probability also decreases with the total loan volume, the drawn percentage in the lender limit, and the percentage of collaterals and securities in the outstanding amount. In summary, our study shows that both accounting-related and loan-related information can help predict more accurately whether a company is likely to be cured following default or not. Additional analyses confirm that our empirical results are robust and not affected by the different settlement periods associated with cured and non-cured companies.

The database we used is a limitation of this study as the study uses pool data on defaulted companies. The database of a single financial institution often contains only a small number of defaulted companies, which leads to an insufficient sample size for estimating the LGD or specific dimensions of the LGD, such as the probability of defaulted companies being cured. A potential solution of this problem is to derive the LGD from the market valuation of credit

default swaps (Baixauli & Alvarez, 2012). Another option is to pool data on defaulted companies such as the member banks of Global Credit Data apply it. As a result, data pooling allows statistically robust estimations on the basis of a reasonable sample size that includes information on the directly observed returns of defaulted companies. However, the validity of pooled data always depends on the comparability of the financial institutions that participate in pooling the data. If the participating financial institutions prepare the data in a different way or do not have comparable internal processes in dealing with defaulted companies, the validity of estimations based on pooled data may be distorted. The potential heterogeneity of the pooled data is a limitation of our empirical analysis. However, we have to acknowledge that the Global Credit Data database is a reliable source of data on LGD and PD.

Another issue is the effect of a bank's workout policy on the independent variables that the empirical analysis takes into account. Due to the lack of data we cannot control for bank-specific workout policies by applying generalized mixed models. At present, 55 international banks pool their data on defaulted loans in the Global Credit Data database. Therefore, the workout policy of a single bank has only a small effect on the entire Global Credit Data database. Although it is very unlikely that international banks execute very different workout policies and simultaneously show very different loan portfolio compositions, an independent variable that captures bank-specific workout policies could clarify whether and to what extent a bank-specific workout policy affects the probability of a defaulted company being cured.

Furthermore, we acknowledge that, although our study is based on a large and comprehensive data set obtained from Global Credit Data, our selection of independent variables may be incomplete. For example, it is possible that we have omitted certain independent variables that might capture a defaulted company's assets, financial situation, and business model in greater detail. We suggest that introducing additional accounting-based independent variables commonly used in bankruptcy prediction models could further increase the validity of the models we applied. Notwithstanding this possibility, it is essential to use accounting-based information, which provides detailed insights into an indebted company, in order to accurately predict a company's probability of being cured following default. In turn, this probability can help more accurately predict the LGD of companies that are cured after default.

REFERENCES

Altman, E. I., & Saunders, A. (1998). Credit risk measurement: Developments over the last 20 years. *Journal of Banking and Finance*, *21*, 1721–1742. https://doi.org/10.1016/S0378-4266(97)00036-8

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature*, *19*, 483–536.

Baixauli, J. S., & Alvarez, S. (2012). Implied severity density estimation: An extended semiparametric method to compute credit value at risk. *Computational Economics*, *40*, 115–129. https://doi.org/10.1007/s10614-011-9290-y

Balcaen, S., & Ooghe, H. (2006). 35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems. *British Accounting Review*, *38*, 63–93. https://doi.org/10.1016/j.bar.2005.09.001

Baruch, L., & Sunder, S. (1979). Methodological issues in the use of financial ratios. *Journal of Accounting and Economics*, *1*, 187–210. https://doi.org/10.1016/0165-4101(79)90007-7

Basel Committee on Banking Supervision. (2019). *CRE—Calculation of RWA for credit risk. CRE36—IRB approach: minimum requirements to use IRB approach*. Basel: Bank for International Settlements. Retrieved from https://www.bis.org/basel_framework/chapter/CRE/36.htm?inforce=20220101

Bastos, J. A. (2010). Forecasting bank loans loss-given-default. *Journal of Banking and Finance*, *34*, 2510–2517. https://doi.org/10.1016/j.jbankfin.2010.04.011

Beaver, W. H., Correia, M., & McNichols, M. F. (2010). Financial statement analysis and the prediction of financial distress. *Foundations and Trends in Accounting*, *5*, 99–173. https://doi.org/10.1561/1400000018

Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education*, 33, 1–43.

Beran, J., & Djaidja, A. Y. K. (2007). Credit risk modeling based on survival analysis with immunes. *Statistical Methodology*, 4, 251–276. https://doi.org/10.1016/j.stamet.2006.09.001

Berg, D. (2007). Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry*, 23, 129–143. https://doi.org/10.1002/asmb.658

Calabrese, R. (2014). Downturn loss given default: Mixture distribution estimation. *European Journal of Operational Research*, 237, 271–277. https://doi.org/10.1016/j.ejor.2014.01.043

Calabrese, R., & Zenga, M. (2010). Bank loan recovery rates: Measuring and nonparametric density estimation. *Journal of Banking and Finance*, 34, 903–911. https://doi.org/10.1016/j.jbankfin.2009.10.001

Chava, S., & Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance*, 8, 537–569. https://doi.org/10.1093/rof/8.4.537

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15–18. https://doi.org/10.2307/1268249

Couderc, F., Renault, O., & Scaillet, O. (2008). Business and financial indicators: What are the determinants of default probability changes? In N. & Wagner (Eds.), *Credit risk: Models, derivatives, and management* (pp. 235–268). Boca Raton, FL: Chapman & Hall/CRC.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44, 837–845. https://doi.org/10.2307/2531595

Dirick, L., Claeskens, G., & Baesens, B. (2015). An Akaike information criterion for multiple event mixture cure models. *European Journal of Operational Research*, 241, 449–457. https://doi.org/10.1016/j.ejor.2014.08.038

Dirick, L., Claeskens, G., & Baesens, B. (2017). Time to default in credit scoring using survival analysis: A benchmark study. *Journal of the Operational Research Society*, 68, 652–665. https://doi.org/10.1057/s41274-016-0128-9

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81, 709–721. https://doi.org/10.1080/01621459.1986.10478327

European Banking Authority. (2016). *Final Report: Draft Regulatory Technical Standards on the materiality threshold for credit obligations past due under Article 178 of Regulation (EU) No. 575/2013* (Report No. EBA/RTS/2016/06). Retrieved from https://bit.ly/3feYghU

European Banking Authority. (2017). *Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures* (Report No. EBA/GL/2017/16). Retrieved from https://bit.ly/321vRIi

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models* (2nd ed.). New York, NY: Springer.

Hagmann, M., Renault, O., & Scaillet, O. (2005). Estimation of recovery rate densities: Non-parametric and semi-parametric approaches versus industry practice. In A. Resti, A. Sironi & E. Altman (Eds.), *Recovery risk: The next challenge in credit risk management* (pp. 323–346). London: Risk Publications.

Hair, J., Anderson, R., Tatham, R., & Black, W. (1998). *Multivariate data analysis* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hosmer, D. W. Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Jones, S. (2017). Corporate bankruptcy prediction: A high dimensional analysis. *Review of Accounting Studies*, 22, 1366–1422. https://doi.org/10.1007/s11142-017-9407-1

Krüger, S., & Rösch, D. (2017). Downturn LGD modeling using quantile regression. *Journal of Banking and Finance*, 79, 42–56. https://doi.org/10.1016/j.jbankfin.2017.03.001

Lev, B. (1969). Industry averages as targets for financial ratios. *Journal of Accounting Research*, 7, 290–299. https://doi.org/10.2307/2489971

Lohmann, C., & Ohliger, T. (2017). Nonlinear relationships and their effect on the bankruptcy prediction. *Schmalenbach Business Review*, 18, 261–287. https://doi.org/10.1007/s41464-017-0034-y

Loterman, G., Brown, I., Mertens, D., Mues, C., & Baesens, B. (2012). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting*, 28, 161–170. https://doi.org/10.1016/j.ijforecast.2011.01.006

Maddala, G. S. (1984). *Limited-dependent and qualitative variables in econometrics* (reprinted ed.). Cambridge: Cambridge University Press.

Mo, L. S. F., & Yau, K. K. W. (2010). Survival mixture model for credit risk analysis. *Asia-Pacific Journal of Risk and Insurance*, *4*, 1–18. https://doi.org/10.2202/2153-3792.1061

Nelder, J. A., & Lee, Y. (1998). Joint modeling of mean and dispersion. *Technometrics*, *40*, 168–171. https://doi.org/10.1080/00401706.1998.10485225

Porath, D. (2006). Estimating probabilities of default for German savings banks and credit cooperatives. *Schmalenbach Business Review*, *58*, 214–233. https://doi.org/10.1007/BF03396732

Rauhmeier, R. (2011). PD-validation: Experience form banking practice. In B. Engelmann & R. Rauhmeier (Eds.), *The Basel II risk parameters: Estimation, validation, stress testing — with applications to loan risk management* (2nd ed., pp. 311–347). Berlin: Springer. https://doi.org/10.1007/978-3-642-16114-8_15

Renault, O., & Scaillet, O. (2004). On the way to recovery: A nonparametric bias free estimation of recovery rate densities. *Journal of Banking and Finance*, *28*, 2915–2931. https://doi.org/10.1016/j.jbankfin.2003.10.018

S&P Global Ratings. (2019). *Default, transition, and recovery: 2018 annual global corporate default and rating transition study*. Retrieved from https://www.spratings.com/documents/20184/774196/2018AnnualGlobal CorporateDefaultAndRatingTransitionStudy.pdf

Sheather, S. (2009). *A modern approach to regression with R*. New York, NY: Springer.

Studenmund, A. H. (2016). *Using econometrics: A practical guide* (7th ed.). Harlow: Pearson Education.

Tong, E. N. C., Mues, C., & Thomas, L. C. (2012). Mixture cure models in credit scoring: If and when borrowers default. *European Journal of Operational Research*, *218*, 132–139. https://doi.org/10.1016/j.ejor.2011.10.007

Wagner, H. (2016). Default definition under Basel. In N. Siddiqi (Ed.), *Intelligent credit scoring: Building and implementing better credit risk scorecards* (2nd ed., pp. 119–130). Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9781119282396.ch7

Wilson, J. R., & Lorenz, K. A. (2015). *Modeling binary correlated responses using SAS, SPSS and R*. Cham: Springer. https://doi.org/10.1007/978-3-319-23805-0_12

Wolter, M., & Rösch, D. (2014). Cure events in default prediction. *European Journal of Operational Research*, *238*, 846–857. https://doi.org/10.1016/j.ejor.2014.04.046

Yildirim, V. (2008). Estimating default probabilities of CMBS loans with clustering and heavy censoring. *Journal of Real Estate Financial Economics*, *37*, 93–111. https://doi.org/10.1007/s11146-007-9046-6

Zhang, N., Yang, Q., Kelleher, A., & Si, W. (2019). A new mixture cure model under competing risks to score online consumer loans. *Quantitative Finance*, *19*, 1243–1253. https://doi.org/10.1080/14697688.2018.1552791